

# Homologia persistente, machine learning e o estudo de proteínas

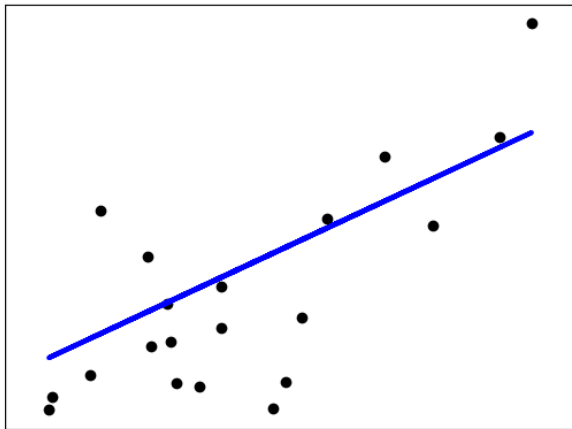
Carlos Ronchi

25 de outubro de 2018

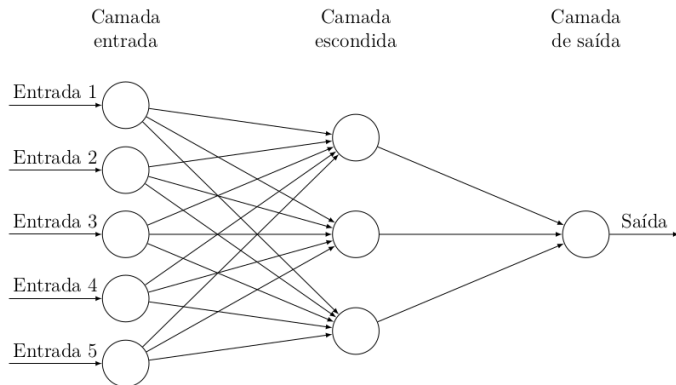
# PREVISÃO

Como podemos usar o computador  
para prever situações?

# REGRESSÃO LINEAR



# REDES NEURAIS



# EXEMPLO

Seja  $\theta$  um vetor que representa os valores  $W_1, b_1, W_2, b_2$ .

$$h(\theta; x) = g_2(W_2 * g_1(W_1 * x + b_1) + b_2),$$

onde  $g_1, g_2$  podem ser funções identidade, sigmoideal e arctan.

$$\min_{\theta} \frac{1}{2n} \sum_{j=1}^n \|h(\theta; x_j) - y_j\|^2$$

# TEOREMA DA APROXIMAÇÃO UNIVERSAL

## TEOREMA

*Seja  $g(\cdot)$  uma função não constante, limitada, monotonicamente crescente e contínua. Seja  $U \in \mathbb{R}^m$  um conjunto compacto qualquer. O espaço das funções contínuas em  $V$  é denotado por  $C(V)$ . Então, dada qualquer função em  $f \in C(V)$  e  $\epsilon > 0$ , existe um inteiro  $N$ , constantes reais  $v_i, b_i \in \mathbb{R}$  e vetores reais  $w_i \in \mathbb{R}^m$ , com  $i = 1, \dots, N$  tais que podemos definir*

$$F(x) = \sum_{i=1}^N v_i g(w_i^T x + b_i)$$

*como uma aproximação da função  $f$ , onde  $f$  é independente de  $g$ , ou seja,*

$$|F(x) - f(x)| < \epsilon,$$

*para todo  $x \in V$ . Em outras palavras, funções da forma  $F(x)$  são densas em  $C(V)$ .*

# REDES NEURAIIS CONVOLUCIONAIS

## Convolutional Neural Networks

# TOPOLOGYNET

RESEARCH ARTICLE

## TopologyNet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions

**Zixuan Cang<sup>1</sup>, Guo-Wei Wei<sup>1,2,3\*</sup>**

**1** Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA, **2** Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA, **3** Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824, USA



## Protein-ligand binding affinity

Mede a força ou a tendência da ligação. Quanto maior o valor, maior a força da ligação.



# Como prever a afinidade?

# ESCOLHA DE PROTEÍNAS

Proteínas: C,N,O,S;

Ligantes: C,N,O,S,P,F,Cl, Br e I.

## VETORES PARA O ALGORITMO

Seja  $[0, L]$  o intervalo de filtração e  $n \in \mathbb{N}$ .

$$V_i^b = \|\{(b_j, d_j) \in \mathbb{B}(\alpha, C, D) \mid (i-1)\frac{L}{n} \leq b_j \leq i\frac{L}{n}\}\|, 1 \leq i < n \quad (1)$$

$$V_i^d = \|\{(b_j, d_j) \in \mathbb{B}(\alpha, C, D) \mid (i-1)\frac{L}{n} \leq d_j \leq i\frac{L}{n}\}\|, 1 \leq i < n \quad (2)$$

$$V_i^p = \|\{(b_j, d_j) \in \mathbb{B}(\alpha, C, D) \mid (i-1)\frac{L}{n} \geq b_j, i\frac{L}{n} \leq d_j\}\|, 1 \leq i \leq n \quad (3)$$

- ▶  $\alpha$ : seleção dos átomos
- ▶  $C$ : tipo do complexo simplicial
- ▶  $D$ : indica a dimensão do diagrama de persistência

# DISTÂNCIA DE OPOSIÇÃO

$$d^{op}(a_i, a_j) = \begin{cases} d(a_i, a_j) & , A(a_i) \neq A(a_j) \\ \infty & , A(a_i) = A(a_j) \end{cases}$$

Figura: Distância de oposição entre dois átomos

# REPRESENTAÇÕES TOPOLÓGICAS

Set	Atoms used	Distance	Complex	Dimension
1	$\{a \in \mathbb{P}   T(a) = e_p\} \cup \{a \in \mathbb{L}   T(a) = e_l\}, e_p \in \mathbb{P}^*, e_l \in \mathbb{L}^*$	$\sigma^{\mathbb{P}\mathbb{P}}$	-	0
2	$\{a \in \mathbb{P}   T(a) \in \mathbb{P}^*\}$	Euclidean	Alpha	1,2
3	$\{a \in \mathbb{P}   T(a) \in \mathbb{P}^*\} \cup \{a \in \mathbb{L}   T(a) \in \mathbb{L}^*\}$	Euclidean	Alpha	1,2
4	$\{a \in \mathbb{P}   T(a) = C\}$	Euclidean	Alpha	1,2
5	$\{a \in \mathbb{P}   T(a) = C\} \cup \{a \in \mathbb{L}   T(a) = C\}$	Euclidean	Alpha	1,2

# RESULTADOS

**Table 1. Performance comparisons of TNet-BP and other methods.**

Method	$R_P$	RMSE
TNet-BP	0.826 <sup>a</sup>	1.37
RF::VinaElem	0.803	1.42
RF::Vina	0.739	1.61
Cyscore	0.660	1.79
X-Score::HMScore	0.644	1.83
MLR::Vina	0.622	1.87
HYDE2.0::HbondsHydrophobic	0.620	1.89
DrugScore	0.569	1.96
SYBYL::ChemScore	0.555	1.98
AutoDock Vina	0.554	1.99
DS::PLP1	0.545	2.00
GOLD::ASP	0.534	2.02
SYBYL::G-Score	0.492	2.08
DS::LUDI3	0.487	2.09
DS::LigScore2	0.464	2.12
GlideScore-XP	0.457	2.14
DS::PMF	0.445	2.14
GOLD::ChemScore	0.441	2.15
PHOENIX	0.616	2.16
SYBYL::D-Score	0.392	2.19
DS::Jain	0.316	2.24
IMP::RankScore	0.322	2.25





Zixuan Cang e Guo-Wei Wei. “TopologyNet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions”. Em: *PLOS Computational Biology* 13.7 (jul. de 2017). Ed. por Roland L. Dunbrack, e1005690. DOI:

10.1371/journal.pcbi.1005690. URL: <https://doi.org/10.1371/journal.pcbi.1005690>.



Cybenko G. “Approximation by superpositions of a sigmoidal function”. Em: *Math. Control Signal Systems* (1989).

# LINKS

Redes Neurais Convolucionais  
TCC (Machine learning)